confidence interval (CI)
An interval estimate (→ *estimator*) that specifies a range of values so that there is a given *probability* for an unknown parameter to lie within that range. The given probability is known as the *level of confidence*, c. A confidence interval is specified as

$$P(a \leq X \leq b) = c$$

i.e. there is a probability of c for the *random variable* X to take a value between (and including) a and b, where X follows the *probability distribution* of the unknown parameter. Typical values of c are 0.9, 0.95, 0.99, and 0.999. With increasing levels of confidence, the range of the confidence interval also increases.

Due to the *central limit theorem*, the *random variable* $\bar{X}$, which is the *sampling distribution of the mean*, is normally distributed (→ *normal distribution*). So when the unknown parameter is μ, the *population mean*, and $X = \bar{X}$, then X can be normalized (→ *standard normal distribution*), thereby transforming the above formula to

$$P(-z \leq Z \leq z) \; = \; P(\frac{a - \hat{\mu}}{\sigma} \leq Z \leq \frac{b - \hat{\mu}}{\sigma}) \; = \; c$$

At this point, the *quantile function* F_Z^{-1} of Z can be used to find the values for $\frac{a - \hat{\mu}}{\sigma}$ and $\frac{b - \hat{\mu}}{\sigma}$ and hence for a and b:

$$z \; = \; F_Z^{-1}(c + \frac{\alpha}{2})$$

Because the normal distribution is a *two-tailed* distribution, half the size of the *critical region* has to be added to the confidence level c (where $\alpha = 1 - c$). Once z is known, the confidence interval can be specified as

$$P(-z\sigma + \hat{\mu} \leq X \leq z\sigma + \hat{\mu}) \; = \; c$$

The meaning of this statement is that there is a $\alpha = 1 - c$ probability for the parameter μ to lie outside of the given interval while the sample X was skewed by chance. See also: *hypothesis test*, *significance*, *level of significance*.

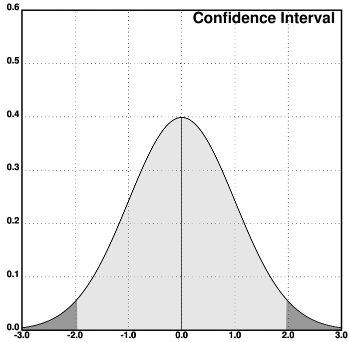

Figure **CFI:** confidence interval on the standard normal distribution with confidence level $c = 0.95$; light gray area: (probability of the) confidence interval; dark gray area: (probability of the) critical regions; significance level $\alpha = 0.05$, two critical regions with probability $\frac{\alpha}{2} = 0.025$.

Example: Given a sampling distribution of the mean $\bar{X}$ of diameters of marbles with estimated mean $\hat{\mu} = 12.5mm$ and $\sigma = 0.2mm$, the confidence interval for the *mean* at a $c = 0.95$ level of confidence would be computed as follows:

$$P(-z \leq Z \leq z) \;=\; P(\frac{a - 12.5}{0.2} \leq Z \leq \frac{b - 12.5}{0.2}) \;=\; 0.95$$

$$z \;=\; F_Z^{-1}(0.95 + \frac{0.05}{2}) \;\approx\; 1.96$$

$$P(-1.96 \cdot 0.2 + 12.5 \;\leq\; X \;\leq\; 1.96 \cdot 0.2 + 12.5) \;\approx\; 0.95$$

$$P(12.11 \leq \bar{X} \leq 12.89) \;\approx\; 0.95$$

I.e., the probability for the true mean μ of the diameter to not lie within the given interval (and the sample being skewed by chance) would be $1 - 0.95 = 0.05$. See figure CFI for an illustration.